Kirtas KABIS System

Automated Book Digitization

Digitize, Enrich, Share

Bound Book Digitization System and Processing

Knowledge Trapped

Until recently, books and manuscripts have been the primary media used by researchers, scientists, historians, philosophers, and authors to record their works for reference or reading by others. The book format has been used for more then a thousand years. The number of book volumes in existence today is enormous and the amount of knowledge they contain is vast and invaluable to all of mankind. Yet, access to most of this information is highly constrained.

The existence of many of these books is often unknown by those that could benefit from them. Historians could expand their research, scientists could advance their work to the benefit of millions. There would be a better understanding of cultures that would advance all societies. The knowledge economy would accelerate if only more people knew of the book's existence, its general content, and had easy access to it.

Easy and quick access to the physical books is a challenge even if its existence is known. It is often impossible to access them at the physical level. Some books cannot be viewed because of the possibility of damage from use. Locked away and rarely viewed books are located around the world in over 200,000 public and private libraries, museums, archives and research facilities. It is an issue and a problem that was foreseen centuries ago….

> *"...let us save what remains: not by vaults and locks which fence them from the public eye, but by such a multiplication of copies, as shall place them beyond the reach of accident."*
> *Thomas Jefferson, February 18, 1791*

Knowledge At Risk

The books on the shelves in the great stone buildings are at risk for loss for many reasons. Some of these books are being lost through natural aging due to the acid in the paper. Some are lost because of public use, abuse and theft, and many are lost because of natural disasters like hurricanes, earthquakes, or tsunamis. The political unrest around the world is another condition that has put these valuable books at risk from terrorist acts. How unfortunate to lose these books, often one-of-a-kind, and the knowledge they contain.

While the physical loss cannot always be prevented, protecting and preserving the knowledge in the books can be achieved as part of the book digitization project.

Trends

There are several trends that have inter-relationships and can positively effect the initiatives for book digitization. All of them are advancing independently yet their synergy offers the potential to release and make available all of the world's knowledge!

Globalization - The country's and culture's of the world are becoming less and less isolated. Changes in manufacturing and services delivery are dramatically different then just a decade ago. The economy is becoming more interdependent among countries. Virtual services are available around the world 24 hours a day. No longer is any country totally self-sufficient.

Knowledge - The Knowledge Economy is growing rapidly and those that can access that knowledge will be more successful in their research because they can more easily build on prior works. Others will be more competitive because they know more or can access more knowledge. The knowledge work-force is growing every year and represents a sizable part of the work force.

Digitization - With all the growth in the digital world, more the 85% of the total information is still stored in paper form. Now we can easily and inexpensively digitize that information to make it more readily available for others to access. Digitization also serves as a means to preserve the content of the books. Digitization technology continues to increase in value.

Internet - By sharing the information, the gap between developed and under-developed nations can be narrowed. Cultures will have a better understanding of one another. This technology has been rapidly adopted and allows us to go almost anywhere in the world in seconds. This conduit has expanded from its early government and research use to commercial and personal use that few could imagine 10 years ago. Investments continue to enhance and upgrade it and Internet 2 will be upon us shortly.

Book digitization needs to be done to achieve two important goals.

- release the knowledge from the shelves to help mankind
- preserve the content of the information for future access

The digitization has to be done rapidly because more people are growing up believing that if something isn't on-line, it doesn't exist. If we don't digitize the rare books soon, we may never have the opportunity if they are lost to decay or disaster.

Objective of Digital Libraries

Having digital libraries that contain catalogued and indexed books and documents makes the library's content accessible to anyone with internet access. New search engines can efficiently search through catalogues to find sources and references of the information desired not only with the primary library, but also with any library that is will to share its metadata

catalogues. Making the world's knowledge globally available will add to the educational, economical and scientific advancements to better all people.

Book scanning offers a key part of the solution for releasing the knowledge on the shelves of libraries to the people of the world. Using proven technologies available today, books can be digitized and kept in formats that are easily accessed and read on PCs and mobile devices with internet connectivity. The word content of many of the books can be processed through Optical Character Recognition software (OCR) to create computer based text files that can be loaded into databases. Now the content of the books can be found through word or phrase searching engines provided on the internet or other private programs. This is in addition to the traditional catalogue information that is typically associated with a book.

The image files can be easily produced and kept in multiple file formats to meet different and multiple needs. For example, the files can be stored as TIFF with no-loss compression applied to assure all the detailed information is retained; or they can be converted to a JPEG 2000, or PDF or compressed with other high compression algorithms to allow the files to move over networks; or optimized for internet viewing. Imagine for a moment how wonderful it would be for you to do your research in Paris and easily locate source information in Boston or Moscow or Sao Paulo, to view the full text and all the graphics associated as if the original book was right in front of you!

Preservation of Knowledge Conflict

Physical preservation is one activity to keep knowledge from being lost. However, it can create a conflict for availability of knowledge. Preservation allows the retention of the original material, which will assure future access. The efforts to preserve the physical book are extensive and costly. By preserving the physical book, the content is also preserved. When books are fragile or easily damaged or rare, the tendency is to reduce the access of them (thus reducing wear and tear) by removing them from circulation. The act of preservation usually has consequences of the books being less accessibility in an attempt to prolong their existence. So while the physical book is preserved, the content trapped in the book and is lost or less accessible to the world.

Book scanning facilitates the preservation of the knowledge content within the book (but not the preservation of the physical book). After scanning, the content and knowledge can be made widely accessible whether the book is returned to circulation or not, the content remains accessible. The electronic book is on the server and the physical book is on the shelf or in preservation vault. Having multiple copies of the file on backup media will increase the survival of the file regardless of disasters, decay or theft.

Preservation of the digital file is now easier and less costly than just 5 years ago. Technology obsolescence still occurs and the digital files must be moved forward to the new platforms and technologies of the future. Selecting a mainstream format is important because those formats will be sustained longer; and investments to migrate those formats forward will be larger and achieved earlier due to market demand. The "LOCKSS[1]" principle is simple, and effective, and works easily with electronic files. The conflict of preservation is easily resolved with digitization.

[1] Lots Of Copies Keeps Stuff Safe

Quality and Standards

With the use of any information technology tool, there must be standards established to provide for quality viewing just as there is for interconnectivity and migration. If the quality is poor and viewers cannot obtain the information they desire, then the effort and investment to digitize was wasted. If the images are of poor quality, have obstructions, lack text clarity, or are missing pages, the users will quickly abandon the digital source. An internet visit to some sources of digitally scanned books will reveal the quality level varies widely.

Over time, the standards of quality will evolve and improve from what is the average on the internet today. As the quality standards increase, those files that are marginal will be abandoned or will have to be rescanned to remain "competitive" with other sources. Poor quality images will reflect negatively on the owner.

A primary tenet of book scanning is to scan the book once and have a level of quality that will satisfy the viewers and endure for generations of use. Images must be of very high quality not only for the black and white text, but also for the color graphics content. Some goals that the project should define and achieve are:

- final image set is very high quality that would be equivalent to a file used for print-on-demand operations by a publisher to produce a book to sell to a customer and any derivatives for easy web viewing are a secondary consideration
- flexible enough to handle a wide range of book quality where the pages may be of different weights, color backgrounds, torn, lose, unintentionally covered up, or pages that may be stuck together
- the electronic book file would be formatted to industry standards with technical and structural metadata provided to assure long term preservation of the file
- the scanning and processing will include the metadata for indexing the book in a database repository, aligned with the standards in existence e.g. Dublin Core +2

The file formats chosen should be industry standards and not proprietary. The readers for industry standard formats are readily available and easily installed on users computers. Proprietary formats may prevent users from viewing the images. If access becomes difficult, the files are essentially lost or no longer accessible.

Additional considerations of the system should include:

- OCR processing leading to full text indexing for ease of content searching
- Pagination and structuring metadata (tagging) for easy navigation and use
- Processing automation
- Optimized file size to reduce the storage and network burden.

The primary applications are:

Special Collections	Original Manuscripts
Rare Books and First Editions	Musical Compositions
Brittle Page and Acidified Page books	Land Ownership
Scientific Research Publications	Civil Registries

CAPTURE - Kirtas KABIS and APT BookScan Design Criteria

With these known needs, a product idea was born in Xerox at their Palo Alto Research Center in the 1990's. People with a vision realized the future need to convert books to digital formats at a quality level and productivity rate that exceeded anything in the market at that time. The project team established some key tenets to guide them in their engineering and design efforts.

The engineering design objectives were:

- Book Safety While Scanning
  - Create a non-stressful position for the binding while scanning occurs
  - Create minimal stress on lifting and turning pages
  - Create as flat a page as possible for imaging
- Superior Image Quality
  - Capture pages in full color for accurate representation
  - Capture pages with the superior Modulation Transfer Function available
  - Capture book with full integrity and no page loss
  - Process images to high quality industry standard formats
- Maximize Capture Speed
  - Minimize operator intervention
  - Use high speed capture technology
  - Use robotics for controllable, consistent performance

In late 2003 the first book-safe, high image quality, automated book scanner was shipped by Kirtas – the APT 1200. Since then, more than 250 units of APT's and KABIS's, (its successor) have been installed in fifty countries around the world in libraries and service support organizations, as well as government and corporate departments to scan books in a safe, productive, non-destructive manner.

The Kirtas design is the most successful in the market.

The product reliability is very high
The service incident very low
The achieved throughput is very high

Coupled with the scanner design elements the software component emerged with equally important criteria. Image processing is a reasonably mature technology. The need to include catalogue information from a variety of formats evolved. And the need to create catalogue information presented itself while digitizing special collections, rare book and other object where not metadata existed. The addition of pagination, structuring and tagging brought new requirements, and this information needed to be included in the metadata. Quickly, the metadata requirements for interoperability were demanded. So the international standards promoted by the METS/ALTO specifications and the new Europeana specifications image processing and software needed to be accommodated

The solution needed is not just buying a scanner. The solution is an excellent scanner and excellent software to meet the user's needs and to be efficient and cost effective. The result is the cost per page is the lowest and you have the best return on your investment (ROI).

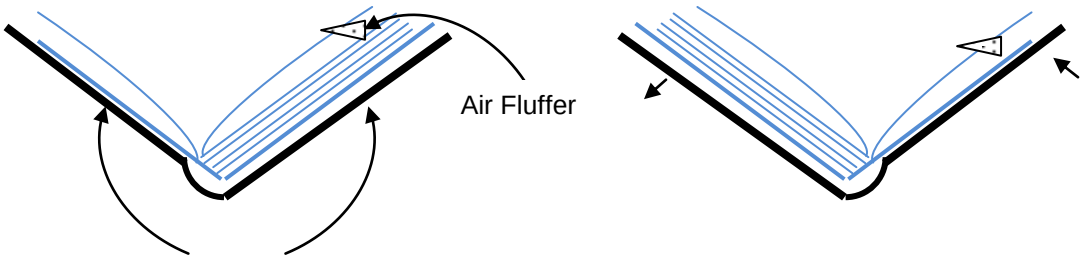
CAPTURE - Non-stressful Binding Position

The engineering design of the KABIS Scanner allows for non-destructive automated book scanning and minimizes the stress applied to the binding and the page turning. Kirtas's two patented technologies are the Smart Cradle™ and the Sure Turn Page Turner™.

The Smart Cradle holds the book open at a non-stressful 110-degree angle. There are sensors that allow the cradle to position the height of the book pages correctly and equally from the imaging camera(s). As the pages are turned from the right side to the left, the cradle technology will automatically raise the right side of the cradle and lower the left side. This keeps the pages in the same position and the same distance from the cameras continuously throughout the scanning of the entire book. .

The next thing that occurs is the action of the air fluffers. The air fluffers placement is on the top and bottom edges of the book on the right side when it is in the cradle. The air fluffers blow air between the top 20-30 pages of the book to help them lift and separate easily. The top page will lift the most. In the operation of the KABIS, this air fluffing occurs many times to achieve the separation. This technique is very gentle on the pages and without stress.

Kirtas KABIS



Articulating sides of cradle
Beginning of book scanning

End of book scanning
Cradle sides move as pages turn to left

Low stress – low curvature

CAPTURE - Minimal Page Turning Stress

The Sure Turn Page Turner includes the vacuum head. It is a design and technology to minimize or eliminate the need for pages to be turned by hand and subject them to damage. The Smart Cradle technology and the Sure Turn Page Turner arm and vacuum head are integrated operationally.

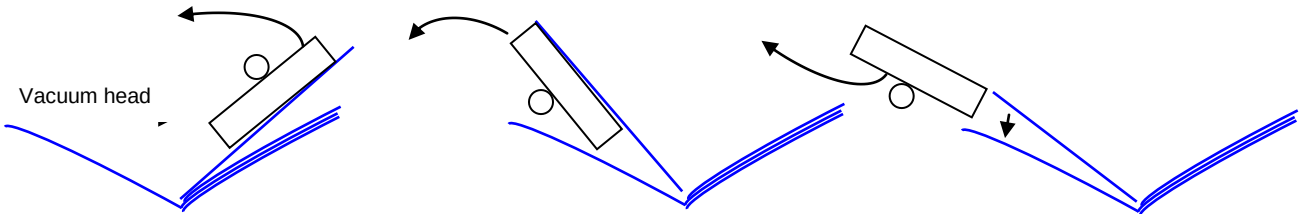
The robotic arm and vacuum head travel to a location determined by the cradle and stops just above the page on the right side. The vacuum lifts the page to the head and the robotic arm moves the head to gently turn the page. As the page moves further to the left, the vacuum head slides out from under the page and allows the page to settle down on the left side.

CAPTURE - Minimal Page Turning Stress

The vacuum has an area of about 18 square inches (116 cm^2) and makes contact with the page at least 0.5 inches from the edges of the page. The page therefore remains nearly flat as it is turned by the movement of the robotic arm on the binding edge. Consequently, the page is turned as gently as possible and the chance of damage is near zero.

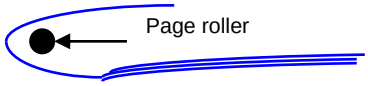
Why? It is around the edges where the pages become brittle. The vacuum head gently lifts the page from the center part of the page and away from the edges. Other devices that require manual page turning by using the hand, or use a mechanical device to cause the page edge to curl to allow separation and turning, or use a roller device to cause page turning will result in brittle page edges to cracking and chipping! It is unavoidable.

Kirtas

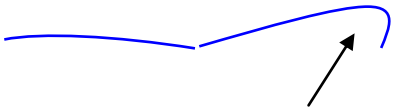


Vacuum head lifts page on right | Head turns page with no bending | Head releases page on left

Other automated and manual page turning products



Severe page bending



Severe page corner bending by hand or mechanical finger

CAPTURE - Robotic Reliability and Page Turning

The robotic arm technology is under system control and will repeat its function in the same way and at the same rate every time. The vacuum head lifts a large area of the page for turning in a gentle manner, not just the corner of a page. Manual page turning for long periods will result in the operator varying greatly the force used and on what part of the page. This will eventually result in operator errors, including page tearing or other damage.

The robotic arm will be gentler on the book pages then the human hand for turning pages, and **the arm will perform at a rate significantly faster than a human can achieve and sustain.**

CAPTURE - Robotic Reliability and Page Turning

Books, pages and bindings vary from book to book. So a system with only one setting will not succeed in safe handling. The KABIS system allows the operator to change the following functions to assure the best performance and reliability

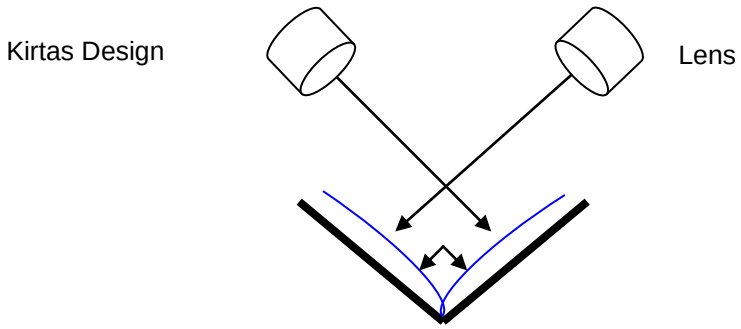
- Speed of the robotic arm. If the paper is very light weight, then going slower is more productive. Some very fragile pages already suffering from brittleness or flaking should be done at slower rates. Normal and heavier weight page paper can operate at high speeds.
- Vacuum lift: The vacuum lift can be controlled to be stronger for thicker, heavier page paper or lowered for very thin page papers
- Air Fluffers: The amount of air needed for fluffing is greater for heavier page paper than for light weight page paper

With these controls, the operator can set the KABIS to the proper optimal levels for the book to assure the quality, safety and productivity.

Again, the consistency of the robotic page turning under system control does not vary as it would with manual page turning.

CAPTURE - High Quality Imaging

The flatter the page is the better for imaging sharpness. With the Kirtas design, each page is intentionally scanned separately. Just by having the book in the cradle, most of the curvature problem is removed. The two clamps, positioned at the edge of the page near the binding, automatically and gently aid in page flattening after each page is turned. These clamps apply only a very gentle pressure to the page without creating stress on the binding. The clamps are easily positioned for varying book sizes and only touch the edge of the page with a downward contact. The depth of focus of the camera lens overcomes the slight remaining curvature (usually less then 0.5 inch). Any remaining page curvature is removed with the technology in the image processing software.



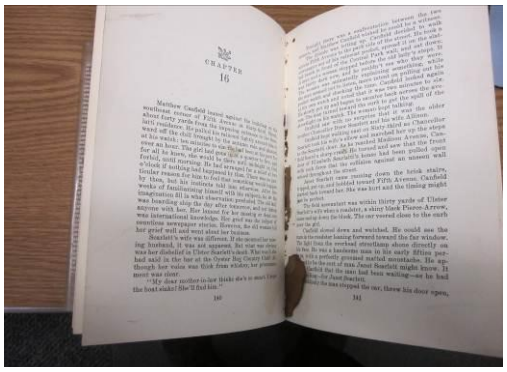
Low stress on page and binding – Curvature minimized

CAPTURE - Fully Automated Solutions to Ignore

A few automated solutions are available and like to ignore the need for operator attendance. "Put the book in, push Start, and come back when completed." While this may work for a very small percentage of the book population, anyone that is knowledgeable about books knows that books can vary substantially. For example:

- bindings are stiff when new or unused, looser when highly used, to ineffective
- some pages may have corner folds
- pages will have different curl characteristics
- paper weight can vary widely
- pages can be stuck together
- loose items are in books that will block the capture of the text or picture behind them

Is it reasonable to think that one setting, operating unattended will accurately scan a whole book from start to end without any stopping?



This is an example of what I found 3 times in a used book I ordered on ebay

These variations above will create different page turning requirements or require a 'Pause, Adjust, Restart" sequence to produce quality images.

CAPTURE - Operator Assistance

Because of these variables, it is beneficial to have a system that will allow the operator to have easy and immediate access to provide assistance to overcome these variables. The result will be a higher productivity with a high quality and fewer rescans. There are three scenarios possible:

- the operator can effect the assistance without stopping the robotic operation
- the operator can quickly and easily stop and restart the robotic function
- the operator can manually turn pages for short runs as needed

For example:

- Stiff bindings: some books will not open easily and page curl can be excessive resulting in poor image quality. the operator can help flatten pages
- Loose pages: the operator can stop the robotic arm and manually turn the pages; then restart once past the loose pages.
- Loose items: the operator can pause the operation, remove the note or debris and quickly restart the operation.

With the Kirtas design, the system will pace the operator. This is significantly different then the manual scanners. With the Kirtas design, the operator can intervene and assist the scanner without slowing down productivity – no other system offers this. With the Kirtas design, one can change between manual and automated scanning with just a touch of a button – no other system offers this. The result is the ability to achieve very high productivity across a wide range of books with different characteristics, yet maintaining the best image quality possible.

The speed of the robotic arm is suitable for a wide range of book variations, yet it is controllable to meet the characteristics of the unusual book.

CAPTURE - Automated High Production Scanning

In order to achieve the highest productive throughput, the system must be indifferent to common variables of scanning. Typically, other systems will slow down significantly as these variables occur. For example, the table below illustrates a comparison between the Kirtas scanner with another scanning system. The number in the column represents the real percentage of throughput achieved compared to the manufacturer's quoted rate for best case scenario. Most manufacturers state the scan rate at 200 dpi, and letter/A4 size documents. Multiply that rate by the number the column to learn what the real scanning productivity number is. Please note that Best Case rate is not the same for each scanner. If Best Rate for Kirtas is 800 pages per hour and the Other scanner is 1000pph, then for 300 dpi scanning the Kirtas system is 800pph and Other system is 500-800pph. In addition, if the 300 dpi scanning is done in color, then the rates are Kirtas 800pph and Other is only 150-400pph!

Variable	Kirtas	Others
Best Case[1] Rate	100%	100%
Resolution to 300dpi rate	100%	50-80%
24/36 bit color rate	100%	30-50%
Grays scale rate	100%	40-60%
8.5" to 14" page size rate	100%	70-80%

[1]Best Case = black & white, 200dpi, 8.5" wide

The cumulative effect of these variables may cause a significant change in quoted scanning speeds from 100 to 20. This variation must be accounted for in the project cost proposal, or one will have errors in forecasting resource needs, completion time and overall costs!

CAPTURE - Page Capture

The system starts with a digital photograph of the page, using a 21.1 megapixel camera. The camera's optical path is designed to look at one page at a time in the 110 degree cradle. The positioning puts the camera at a 90 degree angle to the page. This position minimizes the curvature effect that occurs near the spine. The text on the page is nearly flat in this position, and the depth of field is nominal so sharpness is maximized.

With a digital camera, the digital array is less likely to have imaging defects because the scanner head is stationary and all pixels are sensitized at the same moment. The better the original capture, the better the final image. The camera shutter speed is fast so the page is captured in less then a second compared to 3-6 seconds for linear scanners. This allows the system to be many times faster then others! Real productivity is achievable. And all this is done with full color imagery. All the graphic and picture content in books is faithfully captured for viewing and printing. The KABIS III system uses two cameras; one looks at the left-hand pages and the other the right hand pages. Thus the distortion that can occur from page curl near the binding becomes a non-issue even when text is very close to the binding edge. Each camera is connected to the PC via high speed USB so the capture speed remains continuous without slowdown or buffering required.

CAPTURE - Image Quality

With the Kirtas design, the capture process is under system control improving the results due to the following:

- Image capture is 'still photography' at less then 1/80th of a second. The page is still and flat when the camera captures the page. Nothing is moving. This eliminates image blurriness, and curvature at the spine.
- There is minimal operator intervention during the capture process thus eliminating the presence of hands, arms, etc. in the image.
- The limited operator intervention reduces fatigue that causes other image quality defects.
- Sharper images are generated yielding significantly better OCR processing and text search-able images.

CAPTURE - Book Security and Preservation

With the Kirtas KABIS system, you can take the solution to the books at the library and do the scanning on-site. This means one can scan any rare book in the collection with great care and under security without having to give up possession of the book. The scanning is done with your own people supervising the scanning personnel giving them comfort and confidence in the book handling. Thus, book security is achievable to the level you want to implement.

We have tested and scanned acid paper book pages. The robot is gentler then the human hand and less damaging. Kirtas is successfully scanning these types of books for its customers.

The light bulbs are Phillips T41 compact fluorescent and they have negligible UV.

The KABIS system can be operated in manual mode for books with loose pages or torn pages. This allows the Capture to occur with the same imaging technology as the rest of the book and the collections. While manual page turning is required, the risks are limited just to these small quantities of pages.
The manual mode reduces the productivity rate to about 600 pages per hour.

CAPTURE - Summary

The integration of the technologies is what achieves the results and reliability. It isn't just the cradle or just the robotic arm…..it is the whole design that makes this successful Capture.

ENRICH - Image Processing

When you start the processing with a high quality image, the processing for deskew, despeckle, rotation, etc., naturally yields the best result possible. LIMB software offers an extensive and robust suite of image processing tools that operate and very high speed to quickly produce finish, user-ready images.

During the image processing, the original source image is retained as the potential 'archive image' or in case some reprocessing or selective processing is desired, thus eliminating the need to re-scan. Image processing can be done in a single book batch mode, or multiple book batch mode overnight.

Image Processing Features of LIMB (some, not all)

- Crop image based on mask or border
- Remove clamp image
- Deskew on text
- Page curvature correction
- Image rotation
- Brightness to lighten or darken
- Contrast to bring out detail
- Image sharpening adjustments to create crisp edges on characters to improve OCR
- Image segmentation to allow color retention of pictures and graphics, while text is in black and white
- Cloning to replace speckles or other unwanted artifacts to match the page color
- Pagination to facilitate image numbering to align with page number and to assist in identifying for any missing or duplicate pages
- Automated tagging and structuring from the Table of Contents with links to the chapter's first page
- Automated quality control checking for specific defects and visually tagging the image for operator attention and action
- Retain book covers in original color and size
- Centering of text to the page frame
- Page padding to align the image with the original page size or prepare for reprinting the book
- Checksum to allow for future checking to detect page alteration

All processed images are a derivative from the original raw image file that is retained for subsequent processing or re-evaluations and negating the need for rescans!

ENRICH - Process New Images or Existing Images

LIMB can be used to process images from almost any scanner that is provided in CR2 or TIFF or JPEG file formats. So it is compatible with other scanning devices you may have installed for your digitization project.

In addition, LIMB can be used to re-process existing image files (TIFF; JPEG; JP2, PDF, PNG) to

- improve their quality for viewing and OCR processing; this includes deskewing, curvature correction, sharpening, etc.
- add pagination
- add structuring
- add metadata
- output to additional file formats and image resolutions
- OCR for searchable PDF
- many other functions in LIMB if the image content is available

LIMB brings great features and functions not only to the KABIS scanner images, it adds greater utility to the digitization department and the ability to add more value to existing image files.

ENRICH – Other Key Features

Image Segmentation

One of the key features is image segmentation where the processed full color image is segmented into graphics/pictures and text. The graphic/picture areas are retained in full color (or grayscale) and the text is processed to black and white resulting in the optimized image size without loss of content.

Output Format

One or more of the following file formats can be output by LIMB:

- JPEG or JPEG 2000
- TIFF color uncompressed or LZW, TIFF grayscale or TIFF bitonal G4
- PDF image only or PDF-A, single or multiple pages; searchable PDF
- PNG
-

Metadata

The LIMB software provides Technical, Structural and Descriptive metadata for users wanting this information in *open schema* XML files; data values are preceded by the usual element names (tags).

Technical Data: information about the machine and operation during capture and is automatically generated by the system

Structural Data: information about the book structure/layout requires some manual entry: Table of contents, chapter marks, etc.

Descriptive Data: information about the book that is in MARC, DC, METS, or any OAI database

The descriptive metadata can be entered with Barcode reading of the ISBN or a user interface for entering content descriptions (e.g., Title, Author, Copyright date, Table of Contents, etc.) within LIMB for fast entry and ease of use.

METS ALTO Output Format

Metadata is output in a METS/ALTO compliant file format for use with repository databases or other processing operations.

OCR Processing

LIMB integrates the IRIS software. OCR processing is done after the image processing is completed to achieve the highest yield possible. The OCR output is used to create a text file for indexing, and to produce image-on-text or text-on-image searchable PDF files.

Multiple output options

- Choose the interpolated final resolution desired between 300-600 DPI
- Choose one or more of the image format outputs for the conversion

Automated Table of Contents Tagging/Structuring

LIMB has a uniquely creative method and technology for creating the tagging of images to descriptive information as well as the links from the Table of Contents to the first page of that chapter. In a PDF output, these appear as linked bookmarks. All of this metadata is included in the METS/ALTO output associated with the book

Workflow Queue Monitoring

LIMB presents the status of all the book files in the queue for easy viewing, monitoring and prioritizing for processing activities. The user can easily know what the next activity is.

Book Quality and Image Processing Quality Control

The process starts with a color image to allow processing with all the information

- Regardless of features used, one can always return to the original image for re-processing; re-scanning is very, very rare
- Automated quality control detection for following and the software will visually tag these images for operator review and corrective action if necessary
  - Blurred images
  - Repeating images
  - Text cropped

- Operator has thumbnail images readily available to check processed image quality; easily and quickly correct as needed.

- The image processing software segments text into black and white imagery while retain color for graphics and pictures.
- Metadata is output in METS/ALTO compliant XML format.

ENRICH - Summary

Kirtas is the leader in the market and continues to advance its products, features and productivity to satisfy customer needs with radical, innovative technology.

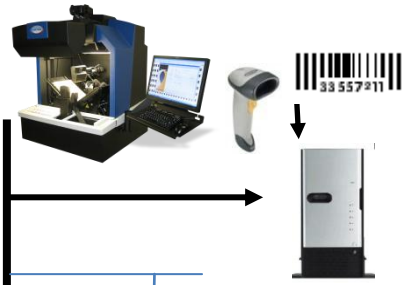
The Kirtas products and process for book digitization offer you the safest and fastest way to produce high quality digital files of books in a non-destructive manner. The process allows you to create the digital book and the necessary metadata about the book to assure easy import and accessibility within a digital book information management system.

You can be sure that the images produced will always reflect positively on the owner. Sometimes you will only have one opportunity to digitize a book, and Kirtas will assure you are successful.

Kirtas Book Digitization with KABIS, LIMB & YOOLIB

Books are delivered; assigned unique barcode and entered into the database.

Capture



Scan barcode to create unique book folder name (ID); Scan book

Capture QC is completed by operator and database is updated

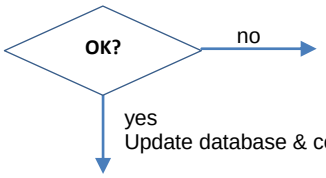


Image Processing

Book Barcode ID
CR2 / TIFF / JPG
and template

LIMB™

Process book images with LIMB and output color, gray, b/w images

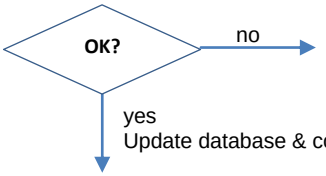
Download MARC, Dublin Core, METS metadata

Quality Control

Book Barcode ID
TIFF / JPEG &
metadata

LIMB™

Processed images are QC'd and corrections made to finish into high quality digital books



abc

ÅŁČ

Conversion

Book Barcode ID
TIFF / JPG / JP2 /
PDF& metadata

IRIS ABBYY Sakhr Software

OPTIONAL with OCR license:: Selected books are OCR'd.

Paginate and create Tagging and Structuring metadata

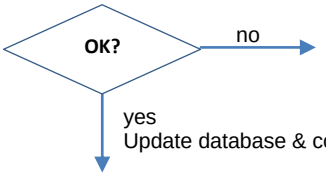
Output METS/ALTO or .csv

Convert images to one or more formats as required.
.tif, .jpg, .jp2, .pdf, .pdf-a, .png
text output in .txt;

Book Barcode ID
Image output &
all metadata

LIMB™

Final QA of books



Publish

YOOLIB

Digital books and metadata are loaded to digital library system

Physical books are re-shelved

Kirtas – From Shelf to Net…As Easy as 1, 2…3